

УДК 811.162.1=162.2'374.822:004.65

**ПОЛЬСЬКО-УКРАЇНСЬКИЙ СЛОВНИК
ЗА РЕДАКЦІЄЮ ЛУКІЇ ГУМЕЦЬКОЇ ЯК ОСНОВА
ДЛЯ ОБ'ЄКТИВНОЇ БАГАТОФУНКЦІОНАЛЬНОЇ
ЛЕКСИКОГРАФІЧНОЇ БАЗИ ДАНИХ***

Наталя КОЦИБА

*Варшавський університет
вул. Краківське Передмістя 26/28, Варшава (Польща)
e-mail: natalia@ibi.uw.edu.pl
Інститут міждисциплінарних досліджень*

У статті описується один із важливих етапів створення польсько-української лексикографічної бази даних на основі оцифрованої версії паперового видання польсько-українського словника за редакцією Л.Гумецької, зокрема робота над нейтралізацією змісту словника з точки зору ідеології. Описано прояви ідеологізації лексики у словнику-джерелі в різних елементах його структури та зміни, які необхідно внести, щоб лексична база якомога об'єктивніше відображала реальність. Окреслено перспективи подальшого розвитку проекту.

Ключові слова: польсько-українська лексикографічна база даних, словникова ідеологія, частотний словник, електронний словник.

У 2004–2009 рр. в Інституті славістики Польської академії наук та Українському мовно-інформаційному фонді Національної академії наук України проводилась робота над спільним науковим проектом польсько-українського електронного словника. Словник створювався передовсім з огляду на потребу такого лексикографічного джерела для широкого кола користувачів. Відсутність великого словника для даної пари мов, особливо електронного, доволі відчутна¹. У результаті цієї роботи загальнодоступним для пошуку в мережі інтернет за адресою <http://www.domeczek.pl/~polukr/pus> став польсько-український словник, що налічує близько 30 тис. найбільш уживаних слів.

Проект цей перспективний і багатоаспектний: словник не лише має суто практичне застосування – за його допомогою можна вирішувати також ряд теоретичних лексикографічних проблем. З огляду на це, робота над лексикографічною базою продовжується в Інституті міждисциплінарних досліджень Варшавського університету. Кінцевою ме-

* Авторка хотіла би висловити вдячність професору Янушу Рігеру за ряд цінних зауважень, які були взяті до уваги під час роботи над текстом статті.

¹ Див. огляд сучасного стану польсько-української лексикографії в: *Kotsyba Natalia, Turska Magdalena. Polsko-ukraińska leksykografia – współczesny stan i perspektywy. Semantyka i konfrontacja językowa*. Т.3. Warszawa, 2006. Деякі аспекти також обговорюються у праці: *Karpluk M. Dwujęzyczne słowniki ukraińsko-polskie ostatniego półwiecza. Prace Filologiczne LI. Warszawa, 2006. S.197–208.*

тою проекту є реляційна лексикографічна база даних, що віддзеркалює в достатньому обсязі сучасну лексику обидвох мов, новітню термінологію та слововживання. База даних передбачає наявність розширеного матеріалу для обидвох мовних напрямів. Крім опрацювання лексичного складу бази даних, ведеться також робота над вдосконаленням опису словникової статті, її логічної структури (так званий парсинг) та послідовного поділу на значення. Планується інтегрувати до лексикографічної системи тлумачні словники та корпуси польської та української мов, а в перспективі – інших слов'янських мов. Паралельно ведеться робота зі стандартизації форматів лексикографічного запису, опрацювання інформаційного комплексу та відповідного програмного забезпечення для автоматизації праці лексикографів. Відтак, польсько-українська електронна лексикографічна база даних буде зберігатися та розповсюджуватись у форматах, узгоджених із міжнародними стандартами.

Робота над електронною мережевою версією словника передбачала насамперед оцифрування² і переведення у форму бази даних паперової версії найбільшого зі створених для польсько-української мовної пари словника за редакцією Лукії Гумецької, що вийшов у 1958 р. у Києві³ (далі – Словник) і на даний час є бібліографічною рідкістю. За даними редакторів, Словник налічує близько 100 000 слів, розподілених у трьох томах.⁴ Це найповніший і найбільш якісний польсько-український словник, цінність котрого як лексикографічного продукту неможливо переоцінити.

Цей етап роботи не обмежувався суто технічною обробкою і оцифруванням даних. Адже виданий півстоліття тому словник не може відображати сучасного стану лексики обидвох мов. Реєстровий склад його є значною мірою застарілий, що підтверджується також вказівкою на використання реєстру так званого Варшавського словника, що виходив протягом перших трьох десятиліть минулого століття: “Головним джерелом реєстру словника був так званий Варшавський словник (Słownik języka polskiego, ułożony pod redakcją Jana Karłowicza, Antoniego Kryńskiego i Władysława Niedźwiedzkiego, t. I–VIII, Warszawa, 1900–1927). Крім цього, були використані такі словники: Nowy słownik języka polskiego pod redakcją prof. T. Lech-Spławińskiego (A–motyczka), 1938–1939; m. Arcta, Słownik ilustrowany języka polskiego, wydanie trzecie. Warszawa, 1929; Słownik francusko-polski i polsko-francuski, opracował P. Kalina, cz. druga, polsko-francuska, Warszawa” (Передмова, V–VI).

Отже, з одного боку, маємо надмір слів та понять, що вже давно вийшли з мовного обігу, а з другого – брак нової лексики. Підбір та опрацювання нових словникових статей є завданням складнішим, аніж вилучення невживаної лексики, і буде проводитися на наступному етапі проекту. Натомість, модернізація вже існуючого реєстру – завдання простіше, окрім того, цей процес можна автоматизувати. При укладанні реєстру автори намагалися дотримуватися наукової об'єктивності, використовуючи новітні статистичні та інші комп'ютерні методи обробки інформації. Для збалансованого представлення лексики вибиралася за частотним критерієм – використовувалися

² Робота над технічною стороною оцифрування проводилася протягом 2005–2009 рр. невеликим колективом фахівців. Деякі аспекти праці описано в: *Kotsyba Natalia, Shevchenko Igor*. General architecture and lexical entry structure of the Polish-Ukrainian electronic dictionary // *Metalanguage and Encoding Scheme Design for Digital Lexicography: MONDILEX Third Open Workshop*. Bratislava, 15–16 April 2009. Текст статті доступний на сайті <http://www.domeczek.pl/~natko>.

³ *Słownik polsko-ukraiński: We dwóch tomach*. Kolegium redakcyjne: A.I. Gęsiorski, T.L. Humecka (redaktor naczelny), M. Kiernycki, M.J. Onyszkiewicz, M.I. Rudnycki. Kijów, 1958.

⁴ Словник було зіскановано і оцифровано без поправок за допомогою програми Fine Reader 8.0 в УМІФ НАНУ. Після цього три томи було вичитано та відкориговано на предмет помилок OCR в орфографії та форматування. Перший том опрацював Ігор Шевченко, другий і третій – Наталя Коциба. Потім текстові файли було переконвертовано до бази даних, яка вимагала подальшої технічної редакції. У технічній редакції версії бази даних також брав участь Роман Тимошук. Змістові поправки здійснювали Ігор Шевченко і Наталя Коциба.

дані корпусу польської мови, розробленого в Інституті основ інформатики Польської академії наук⁵ (далі – Корпус) та статистичні функції програми Полікарп (там само).

При ближчому ознайомленні з текстом Словника виявилось, що якість опису словникових статей не є стовідсотково задовільною і вимагає додаткової уваги лексикографів. Зокрема, він дуже перевантажений з точки зору ідеології і радянської пропаганди⁶, а відтак, не може в такому стані бути нейтральною основою для розвитку польсько-української лексичної бази даних. Ці ознаки проявлялись як на рівні словникового реєстру, тобто включених слів і представлених їх дериваційних дерев, так і на способі опису словникових статей: підборі тематичних та стилістичних позначок, коментарів, змісту та багатства ілюстрацій до слів, самого набору тематичних позначок тощо. З огляду на це, було внесено ряд змін з метою збалансувати реєстровий склад Словника та спосіб представлення лексики з тематичної, стилістичної та дериваційної сторін. Наявна у мережі Інтернет версія вже позбавлена архаїчної, застарілої та ідеологічної лексики на рівнях реєстру, тлумачень і метапозначок. Наукова новизна проведеної роботи полягає в систематичній та підкріпленій автоматизованими методами обробці окремих аспектів польсько-української лексикографії. Актуальність дослідження зумовлена передовсім практичним значенням існування словника в електронній формі, що відображає сучасну лексику української та польської мов.

Словники належать до довідкового жанру, тому по самій своїй суті є джерелом інформації, що сприймається як авторитетна, а отже, здатна впливати на формування системи цінностей та світогляду користувача. Це і було причиною використання їх як інструменту ідеологізації в радянські часи⁷. Однак основною функцією мови все ж є комунікація, а не пропаганда, тому в даний час проводиться ґрунтовний перегляд способів подавання лексикографічної інформації та очищення словників від ідеологічного баласту. Разом з тим, було б помилковим автоматично вилучити зі словникової бази цей ідеологічний наріст, оскільки він становить доволі цікавий матеріал для соціолінгвістичних досліджень. Не будучи безпосереднім свідком створення словника-джерела та аналізуючи його зміст зі значної часової відстані, важко визначити, як саме відбувався процес його ідеологічного наповнення: наскільки він був свідомим і вимушеним з боку самих редакторів, а наскільки це були наслідки втручання цензури чи словник просто відображав тогочасний менталітет загалом⁸. У всякому разі, всі ці та описані нижче явища є частиною історії лексикографії і заслуговують на те, щоб їх зафіксувати. Наші спостереження не претендують на обширний теоретичний аналіз ідеологізації в сфері лексикографії, вони виникли радше як супровідні думки в ході практичної роботи над текстом словника, однак вони можуть мати як практичну, так і теоретичну цінність. Практичний аналіз старої ідеології може також допомогти уникнути подібної небезпеки

⁵ Korpus języka polskiego IPIAN: <http://korpus.pl>.

⁶ Див. також працю, що порівнює ідеологію в “Словнику української мови” за ред. І. Білодіда та “Słowniku języka polskiego” за ред. В. Дорошевського: *Winogradnik Jana. Słownikowa ideologia. Prezentacje prac słuchaczy studium doktoranckiego przy Instytucie Slawistyki PAN, kwiecień 2006. Warszawa, 2006.* (Внутрішній документ Інституту славістики ПАН.)

⁷ Тут доцільно підкреслити, що 1) ми описуємо стан Словника, як він сприймається на даний час і 2) не вникаємо у питання, хто і з якою ціллю подавав (або редагував) лексикографічну інформацію таким, а не іншим чином. Навіть якщо таке бачення реальності й відображало свідомість авторів Словника або було вимушене внутрішньою цензурою, щоб словник міг вийти друком, або це були поправки видавництва, то це теж є певним джерелом інформації про ідеологізацію на рівні держави. Саме в такому, узагальненому значенні ми живимо поняття ідеологізації у цій статті.

⁸ На роздуми наштовхують способи русифікації, обсяг якої у цьому словнику настільки значний, а насиченість просторічними українськими відповідниками та суржигом настільки висока, що її важко трактувати як наслідки цензурування. Не виключено, що вони відображали стан розмовної української мови на час створення Словника, але це питання окремого дослідження.

ідеологізації сьогодні, оскільки певні політично-соціальні тенденції існують у кожному історичному періоді.

Передовсім, варто назвати три основні площини, в яких використовуються **механізми ідеологізації**: реєстр словникових статей, ілюстрації до опису значень та метаінформація (стилістичні та тематичні позначки). Доволі цікавою є третя площина, оскільки вона несе найбільше аксіологічне навантаження. Цей рівень інформації є супутнім, він рідко перебуває у центрі уваги пересічного користувача, котрий найчастіше зацікавлений у тому, щоб зрозуміти значення слова або знайти перекладний еквівалент, тому діє на рівні підсвідомості. Ефект цієї дії можна порівняти до криптореклами. Нижче ми зупиняємося детальніше на кожному з рівнів.

При перегляді **реєстру** словника увагу привертають слова з низькою частотністю вживання та нерівномірний тематичний розподіл лексики з переважанням заідеологізованої. Окремо варто було би зупинитися на відсутній лексиці, наприклад на широкому спектрі релігійної лексики, на позначення реалій Австро-Угорської імперії чи Другої Речі Посполитої, які могли ще бути свіжі у пам'яті носіїв обидвох мов та використовуватися в художній літературі XIX–першої половини XX ст. Залучення цієї лексики могло би відтворити повну картину лексичного стану обидвох мов на час створення словника. Оскільки нашим першочерговим завданням є збалансування наявного матеріалу, цей етап ми залишимо для майбутніх досліджень.

Нижче у статті описані **реєстрові відхилення**, які видно неозброєним оком, коли не вимагається спеціальної методики для доказу їх статусу, а також представлені об'єктивні критерії, котрі вказують на необхідність перегляду доцільності включення перелічених лексем до словника або принаймні коригування їх стилістичного маркування.

Було би неможливо системно представити стилістично-прагматичні особливості всього словникового реєстру як один із аспектів однієї невеликої статті, тому обмежимося декількома прикладами, підібраними на основі дериваційних критеріїв. Варто відзначити як **вживання специфічних морфем**, так і **розбудований словотвір окремих дериваційних основ**. Одним із прикладів перших є так звані “ізми”: *pawłowizm*, *kolektywizm*, *korupcjonizm* (поширення корупції, підкупності, продажності), *akademizm*, *leninizm*, *miczurinizm* (мічуринське вчення). На час створення Словника ця лексика була в обігу, але на даному етапі вона вже неактуальна і недоречна.

Детальніше розглянемо дериваційну групу з негативними префіксами *przeciw-* і *anty-*, так звані “антизми”, наприклад, *przeciwsojalistyczny* – антисоціалістичний; *przeciwreligijny* – антирелігійний; *przeciwrepublikański* – антиреспубліканський; *przeciwżydowski* – антиєврейський та ін. Ця вибірка є прикладом регулярної деривації з сильною негативною конотацією. За даним зразком можна збудувати практично необмежену кількість слів, але підбірка основ деривації для цього типу словотвору у Словнику зовсім не випадкова, а її підтвердженість статистичними критеріями сумнівна.

Для порівняння подаємо вибірку найчастотніших слів із префіксами *anty-* з Корпусу з частотою вживання лексеми (загальна їх кількість – 624):

Прикметник	Частотність
antykorupcyjny	1212
antymonopolowy	1154
antyterrorystyczny	829
antykoncepcyjny	476
antykoncentracyjny	465

Прикметник	Частотність
antynarkotkowy	161
antysubsydjny	140
antyamerykański	136
antyrodziny	134
antypoślizgowy	119

Прикметник	Частотність
antykommunistyczny	422
antydumpingowy	377
antyalkoholowy	364
antysemicki	340
antydopingowy	321
antypolski	313
antywłamaniowy	273
antynflacyjny	267
antyrecesyjny	258
antykryzysowy	243
antynikotynowy	172

Прикметник	Частотність
antyaborcyjny	118
antypaństwowy	113
antyrządowy	109
antyspołeczny	102
antyunijny	97
antykorozyjny	97
antynarodowy	92
antyjwojenny	90
antyrakietowy	86
antyżydowski	77

Список реєстрових одиниць Словника з префіксом *anty-*, що були зафіксовані також у Корпусі, з поданням частотності з Корпусу (7):

Прикметник – переклад	Частотність
antypatyczny антипатичний	29
antyspołeczny антигромадський	102
antyseptyczny <i>med.</i> антисептичний	9
antynomiczny <i>fil., praw.</i> антиномічний	2
antyklerykalny антиклерикальний	70
antyrządowy антиурядовий, протиурядовий	109
antykonstytucyjny антиконституційний	19

Перелік слів зі Словника, які не зафіксовані у Корпусі (31):

*antymilitarny, antymaterialistyczny, antyludowy, antyliberalny, antychole-
ryczny, antyimperialistyczny, antykonstytucyjny, antykomunistyczny, antymilitarystyczny,
antydynastyczny, antynacjonalny, antyhigieniczny, antyfilozoficzny, antypaństwowy, antytetyczny,
antysocjalny, antysocjalistyczny, antyсанитарны, antyrewołucyjny, antyreligijny, antypartyjny,
antymonarchiczny, antynaukowy, antynaturalny, antynarodowy, antynarodowościowy, antyjwojenny,
antymoralny, antyalkoholiczny, antypostępowy.*

Зрозуміло, що реалії початку ХХ і ХХІ ст. є різними, але узагальнюючи, можна зауважити значно більше економічне навантаження “антизмів” сьогодення, тоді як у Словнику вони майже без винятків мають ідеологічне навантаження. Питання вживаності лексики, що входить до словників, з одного боку, та представлення лексики, що вживається, у словниках – з другого, заслуговують на окрему увагу, але подана вище вибірка та порівняння розкладу частот можуть дати загальне уявлення про ідеологічний профіль обидвох лексиконів.

Окрім надміру вибраних словотвірних моделей, у Словнику спостерігається також штучне та неспіввимірне з точки зору сьогодення розширення словотвірних гнізд, у тому числі представлення регулярного словотвору у випадках, коли для структурно подібних реєстрових одиниць аналогічні словотвірні гнізда не представлено.

Приклади розбудованого словотвору окремих дериваційних основ:

miczurinowiec мічурінець.

miczurinowski мічурінський.

partia: partyjność, POP (*Partyjna Organizacja Podstawowa*) skr. первинна партійна організація, і

т.д.

z bolszewizowanie *n.* більшовизація (*докон.*).

z bolszewiczycь *div.* bolszewizować.

kolektyw, *m.* колектив.
 kolektywista, *m.* колективіст.
 kolektywistka, *ż.* колективістка.
 kolektywistyczny колективістичний.
 kolektywizacja, *ż.* колективізація; ~ powszechna суцільна колективізація.
 kolektywizator, *m.* колективізатор.
 kolektywizm, *m.* колективізм.
 kolektywizować колективізувати.
 kolektywnie колективно.
 kolektywność, *ż.* колективність.
 kolektywny колективний.

Деривати виконують подвійну роль – за їх допомогою створюються нові поняття та фіксуються концепти самих основ. Активно використовуються словотвірні моделі осі часу з префіксом *przed-*. Як правило, основою виступає певний ідеологічно важливий концепт, поява якого мала відокремлювати новий світ і нове життя від старого, наприклад, *przedkolchozowy* – доколгоспний. За тим же принципом побудовано стилістичну позначку “*przedrewolucyjny*”, про яку йтиметься нижче.

З метою ідеологізації використовувались й **ілюстрації**. За їх допомогою ідеологізацію можна розглядати щонайменше у двох аспектах: кількісному та якісному.

У першому випадку спостерігаємо надмір прикладів до реєстрового слова. Подані словосполучення не несуть нової лінгвістичної інформації, їх вживання не обґрунтовується жодною ідіоматизацією чи поділом на значення, що вело б до потреби використання різних перекладних еквівалентів. Завданням такого типу ілюстрування є закріплення у свідомості користувача певних реалій, сприймання їх як норми, як “хліба насущного”, а також освоєння їх обидвома мовами. Для прикладу наведемо словникову статтю “*partyjny*”:

“*partyjny*” (той, хто належить до партії): *aktyw* ~ партійний актив, -ву (партактив); *grupa* ~*na* партійна група (партгрупа); *komitet* ~ партійний комітет, -ту (партком, парткомітет); *konferencja* ~*na* партійна конференція (партконференція); *legitymacja* ~*na* партійний квиток (партквиток); партійний працівник, -ка (партпрацівник); *praca* ~*na* партійна робота (партробота); *staż* ~ партійний стаж, -жу (партстаж); *szkola* ~*na* партійна школа (партшкола); *zebranie partyjne* партійні збори, -рів (партзбори); *zjazd* ~ партійний з’їзд, -ду (партз’їзд).

Ще один вартий уваги тип ідеологізації досягається за допомогою підбирання ідеологічно навантажених ілюстрацій до нейтральних слів. Цей метод також спрямований на збільшення питомої ваги лексики пропаганди. Нижче подаємо низку прикладів, у котрих спочатку вказується слово, значення якого пояснюється за допомогою даної ілюстрації, а потім подається сама ілюстрація:

stanowić: *rewolucja proletarska stanowi o losach świata* – пролетарська революція вирішує долю світу;
wieczorowy: *wieczorowy uniwersytet marksizmu-leninizmu* – вечірній університет марксизму-ленінізму;
wstrząsy: *wstrząsy rewolucyjne* – революційні потрясіння (струси, зворушення);
zaczątki: *peren. zaczątki partii rewolucyjnej* – зародок (зачаток) революційної партії;
przerastanie: *przerastanie rewolucji burżuazyjno-demokratycznej w socjalistyczną* – переростання буржуазно-демократичної революції в соціалістичну;
okres: *okres od socjalizmu do komunizmu* – перехідний період від соціалізму до комунізму; *okres rewolucji burżuazyjno-demokratycznej w socjalistyczną* – переростання буржуазно-демократичної революції в соціалістичну;
przesady: *przesady burżuazyjne* – буржуазні пересуди;
wychowanie: *antyreligijne* – антирелігійне виховання.

В останньому прикладі поєднані дві площини ідеологізації – ілюстрації та “анти-слова”.

Ще одним способом словникової пропаганди в ілюстраціях було використання невластивих реалій та штучних ситуацій, які мали на меті децентралізацію обидвох країн, підкреслення їх периферійного становища відносно Москви. Це видно, зокрема, на прикладі міри відстані. Так, кілометри ілюструються відстанями не між Києвом (чи Львовом) та Варшавою (чи Краковом), а між Варшавою і Москвою: *Скільки кілометрів до Москви?* Числівники ілюструються за допомогою рублів, злоті не згадуються в ілюстраціях взагалі. Зараз такі ілюстрації здаються недоречними і і непрактичними.

За допомогою ілюстрацій також вводяться слова та поняття, не властиві даній культурі, що могло мати на меті штучне зближення культур, наприклад “житель Варшавської *області*”, а у будь-якому разі сприймається як контрверсійне⁹.

Переклади у Словнику загалом подані правильно і повно, але і в цій, здавалося б, найбільш об’єктивній ділянці двомовного словника знайшлося місце для “зловживань”. Як уже зазначалося, важко виявити причини надзвичайно великої кількості **русизмів** серед перекладних еквівалентів Словника, тому обмежимося констатацією фактів. Під зірочкою подаємо русизм зі Словника, далі – його український літературний еквівалент, а в дужках – російський літературний відповідник.

- *жарений *кофе → смажена кава (жарений кофе)
- *нуждатися → мати потребу/потребувати (нуждаться)
- *могучість → могутність/міць (могущество)
- *вірówka → мотузка/шнур (веревка)
- *лагер → табір (лагерь)
- міліцейський *участок → ділянка (участок)
- *похожий → подібний (похожий)
- *сахарний → цукровий (сахарный)
- *жарке → печеня (жаркое)
- *гравіровка *печатаей → гравірування печаток (гравировка печатей)
- *покрасити → пофарбувати (покрасить)
- *командировочні → добові/відрядні (командировочные)

Окрім лексичних, спостерігаються також граматичні засоби русифікації, зокрема, часто використовуються дієприкметники активного стану теперішнього часу та абстрактні іменники – назви дії із суфіксом *-ка* замість *-ння*, наприклад *переекзаменовка*. На відміну від лексичних русизмів, граматичні не можна вважати специфічною ознакою Словника, оскільки в часи його створення згадані риси перебували у загальному мовному обігу.

Питання **вживання діалектних слів та суржику** серед перекладних еквівалентів в історичній ретроспективі відтинком у століття не є однозначним, оскільки і мова постійно змінюється загалом, і стилістичний статус слів зокрема. Попри це, не можна не занотувати такі випадки, які сьогодні однозначно потрапляють до цих категорій, та їх лексикографічну прагматику. Імовірно, поява деяких з цих слів у нормативному словнику могла бути спрямована на зниження престижу української мови. Декілька прикладів: *лучче* (*краще*), *фамільярничати*, *засідательство*, *рознос*, *рознощик*, *рознощиця*, *улюблिवий* (*улюблений*), *бульйоновий* (*бульйонний*), *любимиця* (*улюблениця*), *провонялий* (*засмерджений*), *директриса* (*директорка*).

Варто згадати про два аспекти наведення **метаінформації**: стилістичні позначки та поділ на теми, причому друге може частково служити першому. Спостерігається або використання позначок у випадку нейтральних слів, або використання невластивих позначок.

⁹ Щодо штатів як адміністративних одиниць США, графств Великобританії чи ландів Німеччини слово “*область*” не вживалося і в радянські часи.

Йдеться, насамперед, про тематичні групи “церковний” та “релігійний”. Поділ між церковною лексикою та релігійною має певні тенденції, але не завжди є послідовним, та й доцільність і спосіб поділу між ними можна поставити під сумнів, принаймні, піддати ґрунтовному перегляду. Позначками “kościelny”/ “церковний” маркується лише християнська лексика, тоді як позначки “релігійний”/“religijny” вказують на лексику більш абстрактного характеру, ніж “церковна”, наприклад *czcić*, *wybawca*, *pietyzm*, *душа*. Варто зазначити, що останні перелічені слова вже давно ввійшли до загального вжитку, втратили початкову конотацію, і занесення їх до певної тематичної групи можна трактувати як спробу їх маргіналізації¹⁰. Адже двомовний словник не повинен виконувати додатково функції тезауруса, наприклад, слово *ніс* не має позначки “текстильний”, а *niebo* – “метеорологічний”.

Вживання позначок “іст.”, “дорев.”, “заст.” або відсутність їх не завжди є послідовними, порівн.: *stójkowy*, *przest.* “городовий, городовик” (*дорев.*) і *przodownik* П, “нижчий поліційний чин у довоєнній Польщі” (подано без позначки. – Н.К.).

Аналіз словникового матеріалу дає підстави стверджувати, що позначка доревольційності виступає як синонім застарілості, хоча загалом представлені під нею поняття ще занадто свіжі, щоб бути віднесеними до класу історизмів чи архаїзмів. Ще однією функцією цієї позначки є негативна конотація слів, і вживається вона в основному при соціальних поняттях. Можна виділити наступні значеннєві групи з подібними тематичними позначками:

Професії, позиції, звання

babka *ż.* 3) (*дорев.* – сільська акушерка) баба, повитуха.

dworski przedrew. 1) придворний, двірський; *c z e l a d ź* *dworska* двірська челядь, двірня.

felfebel m., *wojs. przedrew.* фельдфебель.

felfebłowski wojs. przedrew. фельдфебельський.

feldmarszałek m., *wojs. przedrew.* фельдмаршал.

feldmarszałkostwo n. 1) *розм.* фельдмаршальський чин; 2) *зб.* фельдмаршал з дружиною (для порівняння: подано без позначки. – Н.К.)

feldmarszałkowski wojs. przedrew. фельдмаршальський.

kamerjunker m., *przedrew.* камер-юнкер.

lejbmedyk m., *przedrew.* лейб-медик.

położna акушерка; повитуха (*дорев.*).

przystaw m., *hist. przedrew.* пристав.

rewirowy m., *przedrew.* околодочний.

terminator m., *dorew.* учень-практикант у ремісника.

terminatorka ż., *dorew.* учениця-практикантка у ремісника.

terminować dorew. вчитися у майстра ремесла.

seminarzysta m. семінарист (*дорев.*).

Соціальний статус

świat dorew. *wielki* ~ вищий світ, вище товариство.

elita, *blm.*, *ż.*, *przest.* (*цім товариства*) еліта.

ziemski 2) *існ.* земський; *dorew.* *obywatel* ~ поміщик; *własność* ~ка земельна власність.

dworzanin m., *przedrew.* придворний.

dziedziców przedrew. поміщицький.

ekscelencja m., *przedrew.*, *dypl.* превосходительство.

¹⁰ Змушують замислитися також такі приклади, як *horoskop*, *m. neren.* “прогноз”; *stawiać ~py* “давати (робити) прогноз”, які можна потрактувати як штучне узагальнення значення, представлення значення слова нейтральним синонімом або гіперонімом.

Соціальні явища, які “викорінено” у радянські часи

cenzor *m.*, *hist.*, *przedrew.* цензор.

cenzorski *hist.*, *przedrew.* цензорський.

cenzora *ż.*, *hist.*, *przedrew.* 1) цензура; ~ *przewensyjna* попередня цензура; 2) *перен. див.* цензуrowany.

cenzuralny *hist. przen.* цензурний.

cenzurować *hist.*, *przedrew.* цензурувати.

łóże *n.* \diamond *dopew. dziecko nieprawego* ~*ża* нешлюбна (позашлюбна) дитина.

notować 1) відзначати, відмічати; *dopew. człowiek* ~*wany* людина, відмічена в поліції.

horoskop *m.* 1) *ісм.* гороскоп.

kabalista 1) *ісм.* кабаліст; 2) *перен. несах.* туманний філософ.

Позначка “дореволюційний”, на відміну від “історичний”, вказувала на те з минулого, що повинно було сприйматися радянською людиною негативно і, як наслідок, викорінюватися з її свідомості. Позначка “burżuazyjny” теж несе негативну конотацію, але здебільшого використовується для існуючих явищ, відокремлюється від радянської реальності хронологічно (примітки типу “у буржуазній Польщі”) або географічно (“в буржуазних країнах”). Ось деякі приклади застосування цієї позначки:

gadziowy 1) гадючий, гадовий; 2) *перен. зневажл.* рептильний; fundusz ~ (в буржуазних країнах) рептильний фонд; pisma ~*we* рептильна (продажна) преса.

harcerz 1) (юний) піонер; 2) (в буржуазній Польщі) бойскаут, скаут.

sfera: wyższe ~*tu* (у буржуазному суспільстві) високі суспільні сфери (кола), вищий світ.

scudzoziemczyk денаціоналізувати; ~*się* денаціоналізуватися (у буржуазному суспільстві).

antyłudowy антинародний; ~*wa* polityka rządów burżuazyjnych антинародна політика буржуазних урядів.

Додатковим, хоча і рідко вживаним методом була штучна ідіоматизація конструкцій, представлення їх як фразеологізмів, наприклад, *wzbierający przyływ rewolucji* – наростаюче піднесення революції. Крім цього, до метаопису вкрадалися русизми, наприклад, “шахматний” (“шахм.”) замість “шаховий”, “офіціальний” (“офіц.”) замість “офіційний”. У першому випадку в правильному варіанті скорочення було б “шах.” а не “шахм.”, таким чином помилка була ще й помножена.

Як відомо, словник повинен відображати реальний світ та мову незалежно від пануючого політичного ладу, бути об’єктивним джерелом знань. Звичайно, описуючи складні абстрактні поняття, лексикограф не раз змушений зайняти певну онтологічну позицію, але її проєкція не буде поширюватися на слова з конкретним побутовим значенням. Працює саме над такими словниковими статтями ми собі поставили за ціль під час **деідеологізації словника**. Етапи цієї роботи охоплюють усі перелічені вище площини словника.

Перший крок до “нейтралізування” словника – **виділення найбільш частовживаної лексики** на підставі об’єктивного стилістично нейтрального мовного джерела та співставлення її з реєстром. Частотний список було складено на основі Корпусу польської мови, опрацьованого в Інституті основ інформатики ПАН та програми Полікарп. І хоча Корпус охоплює 100 мільйонів сегментів, або приблизно 70 мільйонів слів, значна частина реєстру Словника у ньому не відображена. Спостерігалась і зворотна ситуація – наявність значної кількості доволі частотних у Корпусі слів, які не належать до реєстру Словника. Нас у даному випадку цікавив перетин множин цих двох лексиконів та його частотність. За допомогою статистичних даних вдалося автоматично відсіяти значну частину заідеологізованої лексики, оскільки вона або зовсім не потрапила до перетину, або виявилась украй

низькочастотною. У перспективі треба звернути увагу на жанрову зрівноваженість на репрезентативність корпусного джерела.

Значна увага приділялась **дерусифікації** словника, що здійснювалася двома способами: за допомогою частотних списків тлумачної лексики СУМу (багатотомного “Словника української мови” за редакцією І.Білодіда)¹¹ та інтуїції редактора. Слова, що викликали сумнів як можливі русизми, перевірялися за орфографічним словником, наприклад, *козирок*.

Стилістичні позначки “burżuazyjny”, “dorewolucyjny” змінювались на “historyczny”. Позначки “historyczny” отримали слова радянського періоду, що втратили актуальність, наприклад:

pegeer (PGR – Państwowe Gospodarstwo Rolne), *m., roln., hist.* держсільгосп (державне сільське господарство);
 czekista *m., hist.* чекіст;
 czekistowski *hist.* чекістський.

Корекція ілюстрацій здійснювалася силами самих редакторів. При цьому критерій частотності теж брався до уваги, а перевірялася вона за допомогою пошукової системи Google у мережі Інтернет. Таким чином слово *репресії* ілюструється в контексті *сталінські*. Ще декілька прикладів змін в ілюстраціях:

społeczeństwo: społeczeństwo *socjalistyczne* соціалістичне суспільство → *obywatelskie*;
 епока: w ерсе *socializmu* в епоху соціалізму → *średniowiecza*;
 ulga: ulgi kolchozom пільги колгоспам → *studenckie*;
 mieszkaniec: житель *Варшавської області* → *Warszawy*; мазовецького воєводства.

Звичайно, ідеологічний аспект – лише один із багатьох, які необхідно брати до уваги в ході роботи над великим двомовним словником. З точки зору лексикографії він до того ж є далеко не основним, але подальша робота над словником як основним мовним ресурсом не може проводитися об’єктивно, якщо матеріал не представлено стилістично нейтрально та коректно.

Таким чином, у статті представлено **проект** розробки польсько-української лексикографічної бази даних із урахуванням дотихчасових досягнень лексикографії для цієї пари мов та проаналізовано ідеологічний аспект словника за редакцією Л.Гумецької, який є вихідним матеріалом для бази, на різних рівнях його реалізації. Запропоновано методи нейтралізації змісту словника з точки зору ідеології. Для українсько-польської лексикографії таке системне дослідження електронної версії великого паперового словника робиться вперше. Застосування сучасних комп’ютерних методів обробки мови дає змогу розглядати мовні дані під новим кутом зору, висвітлювати їх у новій перспективі. Представлене дослідження – це тільки приклад величезних можливостей, які дає лексикографам комп’ютеризація праці. Нижче даємо перелік наступних завдань, які планується вирішувати за допомогою згаданого інструментарію.

До **теоретичних завдань**, які можна буде вирішувати за допомогою відкоригованої словникової бази даних, належить дослідження кореляції польських й українських лексичних одиниць на рівні як форми, так і значення, включаючи міжмовну омонімію та лексичні засоби вираження граматичних значень. Пріоритетним завданням є також формалізація опису значень, наприклад шляхом представлення семан-

¹¹ Також використовувалась електронна версія українського граматичного словника, описана в: Шевченко І.В., Широков В.А., Рабулець А.Г. Електронний граматический словарь украинского языка. Megaling’2005. Прикладная лингвистика в поиске новых путей. 27 червня–2 липня 2005. Меганом, Крим, Україна, 2005, ст. 124–129.

Широков В.А., Рабулець О.Г., Шевченко І.В., Костишин О.М., Якименко К.М. Інтегрована лексикографічна система “Словники України”, версія 3.1. Київ, 2007. CD-видання.

тичних кореляцій між лексемами, зокрема виділення лексичних функцій згідно з теорією “Смисл–Текст”, що можуть бути допоміжними при автоматичному перекладі. Представлення дериваційних зв’язків між словами теж може внести ряд переваг. Важливим напрямом є робота над послідовністю граматичного опису, узгодженості формату запису із загальноприйнятими міжнародними стандартами та інтегрування зі словниками інших слов’янських мов у перспективі.

В Інституті міждисциплінарних досліджень Варшавського університету проводиться робота над розширенням польсько-українського паралельного корпусу¹² (PolUKR), який зможе використовуватися для вирішення ряду теоретичних і прикладних лінгвістичних завдань. Одним із основних застосувань корпусу є **верифікація, коригування та поповнення** словникової бази польсько-українських еквівалентів. Це може здійснюватися або напівавтоматизовано, за допомогою конкордансів та лінгвістичного аналізу, або за допомогою спеціальних алгоритмів автоматичної екстракції еквівалентів¹³. Планується інтеграція корпусу з лексичною базою даних для опрацювання термінології новітніх галузей, не представлених у паперовій версії, зокрема таких ділянок лексики, як комп’ютерна, бізнес та новітні технології; сучасної технічної, економічної та розмовної лексики. Окрім паралельного корпусу, який вимагатиме текстів із відповідним лексичним наповненням та якістю перекладу, використовуватимуться і порівняльні корпуси. Також братимуться до уваги вже створені термінологічні словники для даної пари мов.

¹² PolUKR – польсько-український паралельний корпус, доступний за адресою: <http://www.domeczek.pl/~polukr>.

¹³ Turska Magdalena, Natalia Kotsyba. Polish-Ukrainian Parallel Corpus and its Possible Applications // Proceedings of the International Conference ‘Practical Applications in Language and Computers’, 7–9 April 2005. Łódź, 2007.

**THE POLISH-UKRAINIAN DICTIONARY EDITED
BY LUKIYA HUMECKA – A BASIS FOR THE OBJECTIVE
POLIFUNCTIONAL LEXICOGRAPHIC DATABASE**

Natalia KOTSYBA

*Warsaw University
26/28, Krakowskie Przedmieście Str., Warsaw (Poland)
e-mail: natalia@ibi.uw.edu.pl
Institute of Interdisciplinary Studies*

The article describes one of the important stages of the process of creating a Polish-Ukrainian lexicographic database on the basis of a digitalized version of the Polish-Ukrainian dictionary ed. by L.Humecka and published in 1958, in particular the work on neutralizing the dictionary content from the ideological point of view. Manifestation of ideology in various components of dictionary structure and changes that we consider necessary to introduce for the lexical database to reflect the reality as objectively as possible are described. Perspectives of the project's development are presented in short.

Key words: Polish-Ukrainian lexicographic database, dictionary ideology, word frequency, electronic dictionary.

Стаття надійшла до редколегії 12.01.2010
Прийнята до друку 29.01.2010